



OPEN

Primary sequence contribution to the optical function of the eye lens

SUBJECT AREAS:

PROTEINS

PROTEIN FUNCTION
PREDICTIONSK. Mahendiran¹, C. Elie², J.-C. Nebel¹, A. Ryan¹ & B. K. Pierscionek¹¹Faculty of Science Engineering and Computing, Kingston University, Penhryn Road, Kingston-upon-Thames, KT1 2EE, UK,²Department of Bioengineering, Nice Sophia Antipolis University Engineering School, Templiers Campus, 06410 Biot, France.

Received

25 February 2014

Accepted

17 April 2014

Published

6 June 2014

Correspondence and requests for materials should be addressed to B.K.P. (b.pierscionek@kingston.ac.uk)

The crystallins have relatively high refractive increments compared to other proteins. The Greek key motif in $\beta\gamma$ -crystallins was compared with that in other proteins, using predictive analysis from a protein database, to see whether this may be related to the refractive increment. Crystallins with Greek keys motifs have significantly higher refractive increments and more salt bridges than other proteins with Greek key domains. Specific amino acid substitutions: lysine and glutamic acid residues are replaced by arginine and aspartic acid, respectively as refractive increment increases. These trends are also seen in S-crystallins suggesting that the primary sequence of crystallins may be specifically enriched with amino acids with appropriate values of refractive increment to meet optical requirements. Comparison of crystallins from five species: two aquatic and three terrestrial shows that the lysine/arginine correlation with refractive increment occurs in all species investigated. This may be linked with formation and maintenance of salt bridges.

The ability of the visual system to detect and interpret information from light is an evolutionary adaptation following an ancient divergence from a primitive light sensing structure that has resulted in species specific systems with varying degrees of sophistication yet remarkably similar configurations^{1–3}. The eye lens is a key refractive element that has been tailored for each organism in order to meet visual demands required for functional needs^{1,3}. It is composed of lens fibre cells that grow in concentric layers over existing tissue, with no concomitant cellular loss; the tissue accrual is a process that begins in utero and continues through life^{1–4}. The fibre cells are differentiated from epithelial cells located under the anterior part of the lens capsule, which is a semi-elastic basement membrane that contains the lens and transfers the forces needed to adjust its shape with changes in focussing power³. During the differentiation process, the structural proteins of the lens, the crystallins, are synthesised in concentrations that vary across the tissue in order to create a refractive index gradient required for optimising image quality³. Cytosolic concentrations can approach ~400 mg/ml in humans^{3,5–8}. In the terminal step of differentiation the nucleus become pyknotic and the cell enters proteostasis, retaining the requisite concentration and mixture of crystallins^{3,5,9}. The lens, therefore, retains within its refractive index gradient, a chronology of ageing with newly synthesised proteins in peripheral fibre cells and proteins produced during gestation in its centre evolved for longevity and for maintenance of optical function^{3,7,8}.

The major function of the lens is to maintain transparency and to provide sufficient refractive power for light to focus on the retina. In terrestrial species, around two-thirds of the refraction occurs at the air/cornea interface with the lens providing the additional refractive power and, in species with sufficiently malleable lenses, the fine-tuning required to adjust lens shape for different viewing distances^{2,3}. In aquatic species, the comparatively high refractive index of water negates the refractive power of the cornea and has resulted in the evolution of lenses with steep index gradients and high refractive index magnitudes required to provide all or most of the refractive power for the aquatic eye^{3,10,11}.

The refractive index is related to protein concentration by the Gladstone-Dale formula^{3,12,13}. This simple linear equation introduces the refractive index increment (dn/dc), which defines how much a given concentration of a protein will contribute to the refractive index^{3,12–15}. Across the crystallin isoforms differences in dn/dc values have been found, with the smallest of the crystallin classes: the γ -crystallins, having the highest dn/dc ^{13,14}. This is also the crystallin class that is found in the core of the lens^{3,16,17} where refractive index is the highest^{3,18,19}.

It is thought that the requirement for longevity, which is of particular importance in the core of the lens that contains the oldest cells, has been met by the Greek key structure, which is known for its thermodynamic stability^{8,20–23} and is a feature of γ -crystallins and found in the larger $\beta\gamma$ -crystallin family^{8,24,25}. Higher order structures as well as the dn/dc of any given protein are influenced by the primary sequence of amino acids. McMeekin et al measured the dn/dc of each amino acid for 589 nm and at 25°C²⁶ taking into account individual amino acid refractivities and partial specific volume. These values were recently used by Zhao et al^{14,15} to compute



the dn/dc values for an array of proteins. Zhao et al^{14,15} reported the enrichment of aromatic residues, associated with high dn/dc, in lenticular crystallins, particularly in the $\beta\gamma$ isoforms.

Whether there is a correlation between the Greek key structure and primary sequences containing amino acid residues with high refractivities in all proteins or whether the Greek key motifs and amino acids with higher refractivities are found in certain crystallin isoforms was not known. This work suggests that $\beta\gamma$ crystallin isoforms have comparatively high dn/dc values compared to other proteins with Greek key structures.

Results

Specific refractive increments of Greek key domains. Figure 1 shows the dn/dc distribution of the $\beta\gamma$ -crystallin dataset. The mean dn/dc of the $\beta\gamma$ -crystallin dataset is 0.1999 ml/g (SD: 0.0036 ml/g). The mean dn/dc of the double Greek key motif of crystallins was found to be 0.2003 ml/g (SD: 0.0035 ml/g). Plotting predicted dn/dc against sequence length for the 292 entries of the $\beta\gamma$ -crystallin dataset shows a general trend to decreasing dn/dc with longer sequences (Figure 2A). These observations may suggest a link between the double Greek key motif and high dn/dc. It is clear from Figure 2B, which highlights the range of sequence lengths between 50 and 300 residues, that there is a predominance of native proteins with lengths of between 170–180 residues, which corresponds to the double Greek key motif.

The dn/dc of representatives of double Greek key domains from a wide range of proteins were analysed to establish if high dn/dc is a consequence of the motif structure. The distribution of dn/dc for representatives of 52 different double Greek key domains²⁷ indicates that Greek keys from $\beta\gamma$ -crystallin proteins have the highest dn/dc values, appearing as an outlier as seen in Figure 3A. The mean dn/dc value for double Greek keys motifs is 0.1908 ml/g which is close to the mean for human proteins, ie 0.1899 ml/g, but substantially lower than that of $\beta\gamma$ -crystallin (0.2009 ml/g). No correlation is seen between either the sequence length or the number of strands of the double Greek key domain and the predicted dn/dc value (Figure 3B and 3C).

Amino acid compositions and refractive increments in crystallin isoforms. In order to investigate the specificity of the $\beta\gamma$ -crystallin double Greek key domain, a subset of the initial dataset was created containing only proteins within the 170–180 residue range. This comprises 116 $\beta\gamma$ -crystallin isoforms with minimum and maximum dn/dc values of 0.1965 and 0.2084 ml/g respectively and a mean of 0.2024 ml/g. Their amino acid compositions show an

inverse correlation between the proportion of arginine and lysine residues and the dn/dc value ($r^2 = -0.85$, $p < 0.0005$): the sum of arginine and lysine residues is constant at ~ 21 but as the dn/dc value increases, the number of lysine residues decreases and that of arginine increases concomitantly (Figure 4A). A similar inverse correlation is seen with glutamic acid being replaced by aspartic acid ($r^2 = -0.70$, $p < 0.0005$) (Figure 4B). In both cases, an increase in charged amino acids with high dn/dc values is at the expense of those with lower dn/dc values, leading to a higher overall dn/dc value for the given protein. A similar analysis was conducted on the S-crystallin family, which are the major protein class in eye lenses of cephalopods (eg. octopi, squid, cuttlefish). As illustrated in Figure 4C and Figure 4D, inverse correlations between the proportion of arginine and lysine residues, and glutamic acid and aspartic acid are also observed in this family of proteins ($r^2 = -0.92$, $p < 0.0005$ and $r^2 = -0.81$, $p < 0.0005$, respectively). Notably, S-crystallins from the octopus, which are demarcated by open circles (Figure 4C), differ in composition from other S-crystallin sequences. This causes the deviations between the curves representing K and R residues but does not affect the strength of the correlation between them.

Interspecies comparison. Amino acid sequences from a range of crystallins with different dn/dc values were compared in five species to examine whether these specific correlations, more lysine and glutamic acid in crystallins with relatively low dn/dc values and more arginine and aspartic acid in crystallins with relatively high dn/dc values, were consistent across species. Comparison was made between two aquatic and three terrestrial species: ranine (*Xenopus laevis*), piscine (*Danio rerio*), murine (*Mus musculus*), bovine (*Bos taurus*), and human (*Homo sapien*) (Table 1). The greatest variation in dn/dc values is seen in the zebrafish (*Danio rerio*) and the least in the human. The results show that a number of lysine residues in sequences of lower dn/dc values were substituted by arginine in sequences with higher dn/dc values (Table 2). There are no such consistent correlations between glutamic acid residues and aspartic acid residues. Rather, Table 2 also shows that between the protein with the lowest dn/dc and those with the highest dn/dc in any given species, there is a consistent substitution of phenylalanine by tyrosine. In contrast with the other set of substitutions, this creates a very slight decrease in dn/dc.

Salt bridge analysis. Analysis of representative $\beta\gamma$ -crystallin structures indicate that mammalian $\beta\gamma$ -crystallins have around 5 salt bridges per domain length (of around 90 residues); this is not

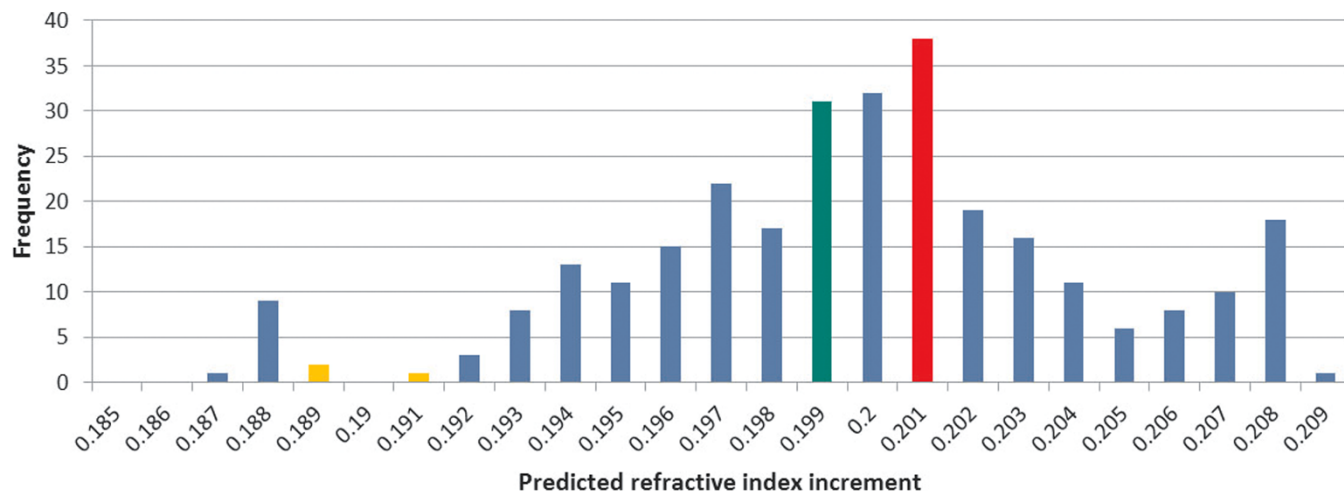


Figure 1 | Frequency distribution of predicted refractive index increments of $\beta\gamma$ crystallins. Green, red and yellow colours indicate the overall average value, the average value of double Greek key motif sequences and the average value of other human proteins, respectively.

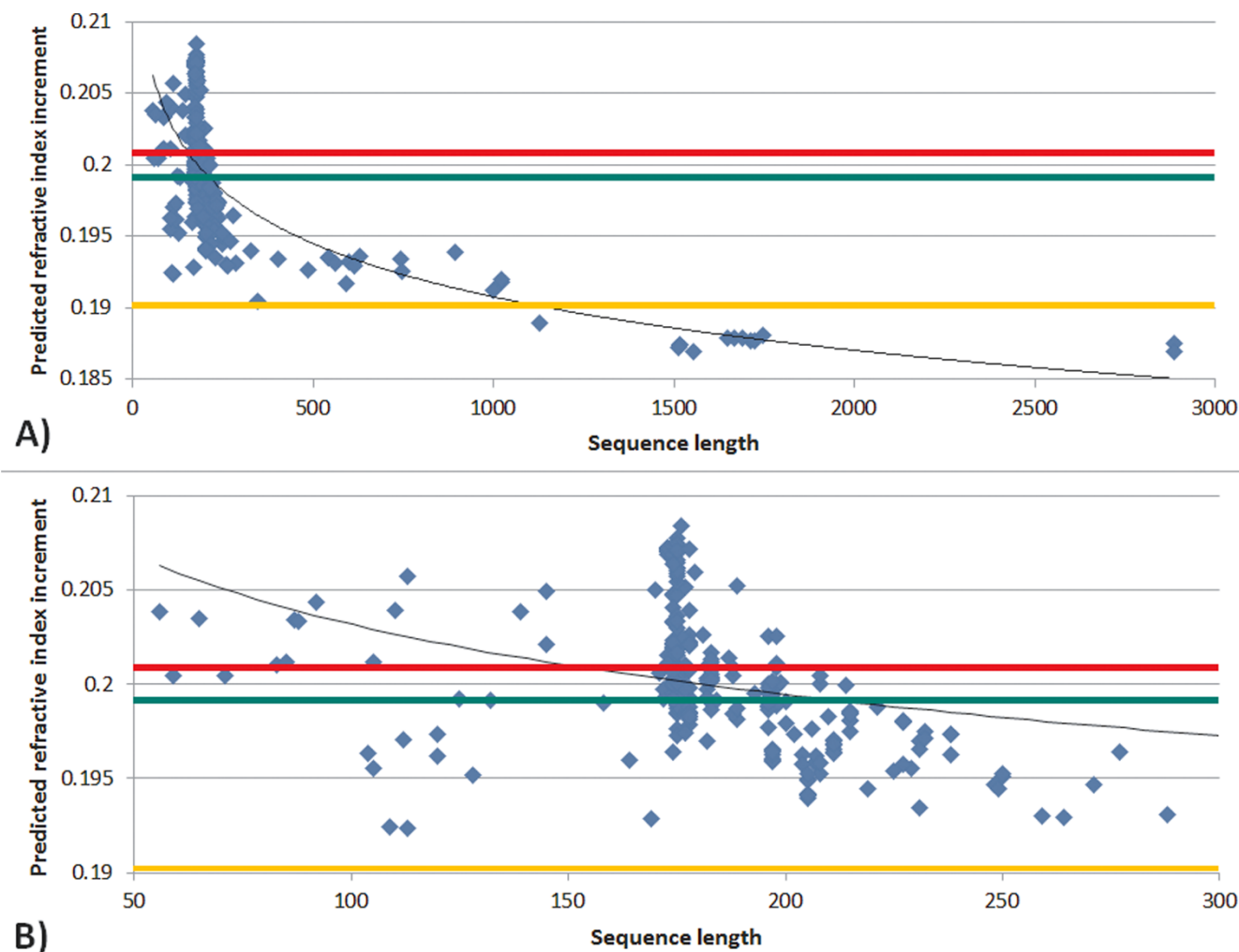


Figure 2 | Predicted refractive index increment plotted against sequence length of $\beta\gamma$ crystallins showing **A)** all crystallins in the dataset; **B)** crystallins the sequence length of which is between 50 and 300 amino acids. Green, red and yellow colours indicate the overall average value, the average value of double Greek key motif sequences and the average value of other human proteins, respectively.

found for the non-mammalian proteins. A separation that is considerably greater than 10 residues is the most frequent separation length (73% of cases) and in cases of separation that are fewer than 10 residues, a separation of 2 residues is the most common (36% of cases); with 21% of cases showing a separation of 7. Although most bridges are formed between amino acids that are >10 residues apart, they are largely from the same domain (84% of cases). Around a third of these cases are interstrand bridges. Figure 5A highlights salt bridge conservation across five $\beta\gamma$ -crystallin isoforms, ie γ B, β B2, β B3, β B1 and β A4. Conserved salt bridges connect two Greek keys within a domain as well as between two domains (Figure 5B and C). The isoforms β B2, β B3, β B1 display two cross domain salt bridges; β A4 has a single salt bridge with a disulphide bond within 4 residues of this. The γ -crystallin (γ B) does not have any cross domain salt bridges; these are prevented from forming because of the negatively charged residue 29 on one domain and corresponding 147 on the other domain. All $\beta\gamma$ -crystallins have a salt bridge between the two Greek keys of the second domain. β A4 also displays a salt bridge between the two Greek keys of the first domain.

A fundamental feature incorporated in the Greek key structure is the conserved β -hairpin motif²⁴. Calculating the values of dn/dc for sequences that comprise the four β -hairpins in each of the isoforms shown in Figure 5A indicates that there is a slight increase in the dn/dc value for these sequences compared to the whole sequence for each respective isoform. The change in dn/dc value ranges from an

increase of 0.36% for β B3 ($dn/dc = 0.1977$ for β -hairpins compared with 0.1970 for the whole sequence) to an increase of 3.37% for β B2 ($dn/dc = 0.2004$ for β -hairpins compared with 0.1939 for the whole sequence).

Discussion

The concept of a refractive increment and the property of contributing to refraction have obvious relevance to the crystallins given the predominant function of the eye lens. It should be remembered that dn/dc is not an immutable property but depends on the solvent²⁸, wavelength²⁹ and, to a much lesser extent, temperature²⁹. Given that experimental crystallin samples prepared for measurement of dn/dc need to be constituted in a solvent that replicates fluid found in the eye lens, the greatest variability in dn/dc values from different studies comes from the wavelength used for measurement. Most experimental studies on dn/dc of crystallins have used proteins from the bovine lens^{13,30–32} and some have concentrated on α -crystallins^{31,32} and/or γ -crystallins³⁰. Where all three broad classes of crystallins have been measured, dn/dc was found to be highest for γ -crystallins¹³. Most importantly experimental studies did not isolate particular isoforms within the crystallin classes. Theoretical studies that have calculated molar refractivities of individual amino acids²⁶ or used these to calculate dn/dc were able to compile these values for many different proteins^{14,15} and provide a deeper insight into the reasons why a protein may have a relatively low or high dn/dc value.

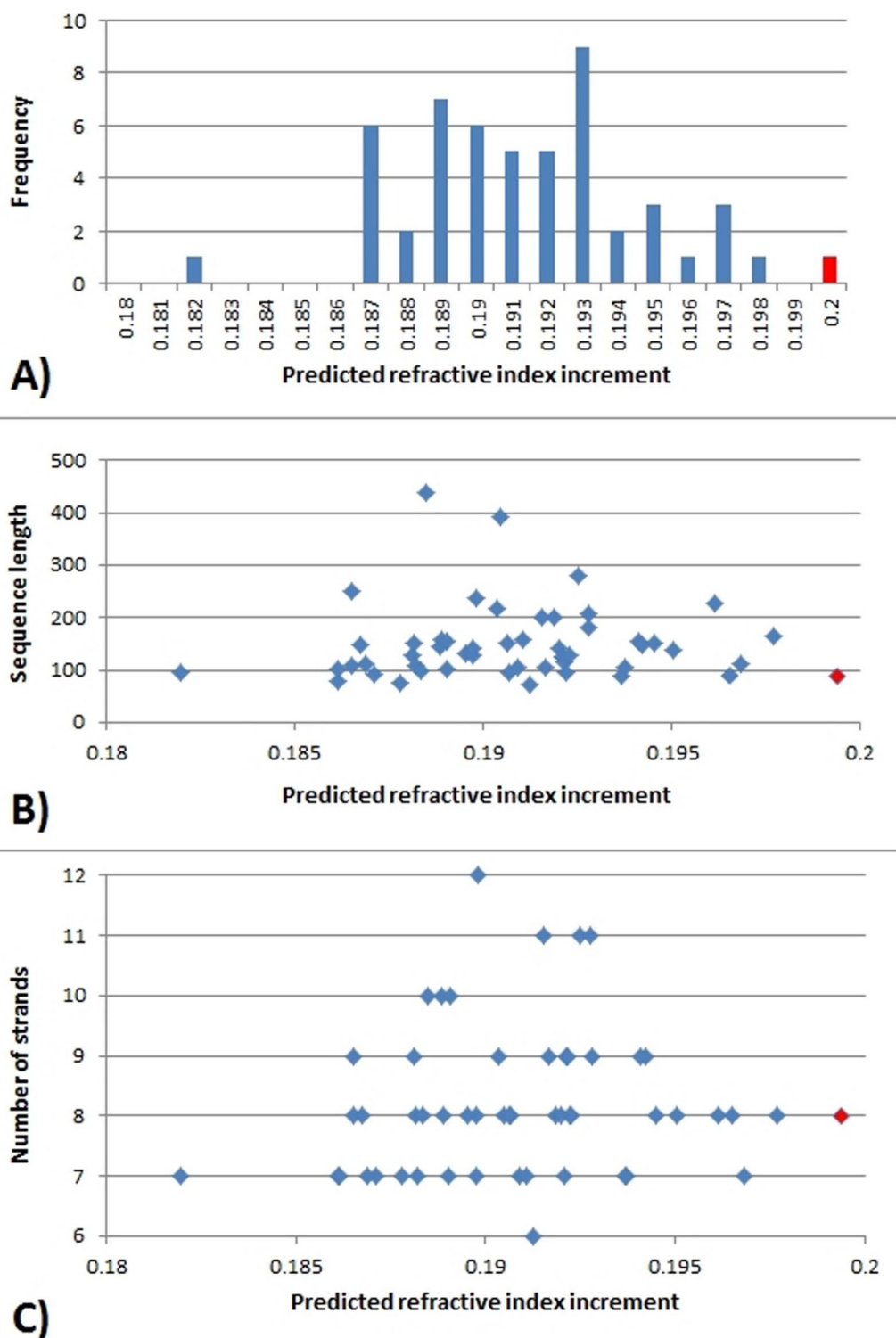


Figure 3 | Frequency distribution A), sequence length B) and number of strands C) plotted against refractive index increments of double Greek key domains. The red colour represents $\beta\gamma$ crystallins.

Aromatic (tyrosine, tryptophan and phenylalanine) as well as sulphur containing amino acids (methionine, cysteine) have relatively high dn/dc values, whilst alanine, proline and serine have the lowest values^{14,15}. The high content of aromatic amino acids in the crystallins^{33,34} coupled with the relatively high cysteine content of γ -crystallins^{24,35–37} provides some explanation for the high dn/dc of this protein class. The γ -crystallin, γM , a protein found in certain aquatic species, has the highest dn/dc (0.209 ml/g) thus far found in a

crystallin resulting partly from its high level of methionine^{14,15}. The high content of methionine may facilitate denser packing³⁵ which would be advantageous for a high refractive index (reviewed in³).

The mean dn/dc of the $\beta\gamma$ -crystallin dataset studied (0.1999 ml/g; SD: 0.0036 ml/g) is significantly higher than the mean dn/dc for other human proteins (mean 0.1899 ml/g; SD: 0.0030 ml/g)^{14,15} and sequence length was found to be inversely correlated to dn/dc . Comparison of proteins with double Greek key motifs shows a sim-

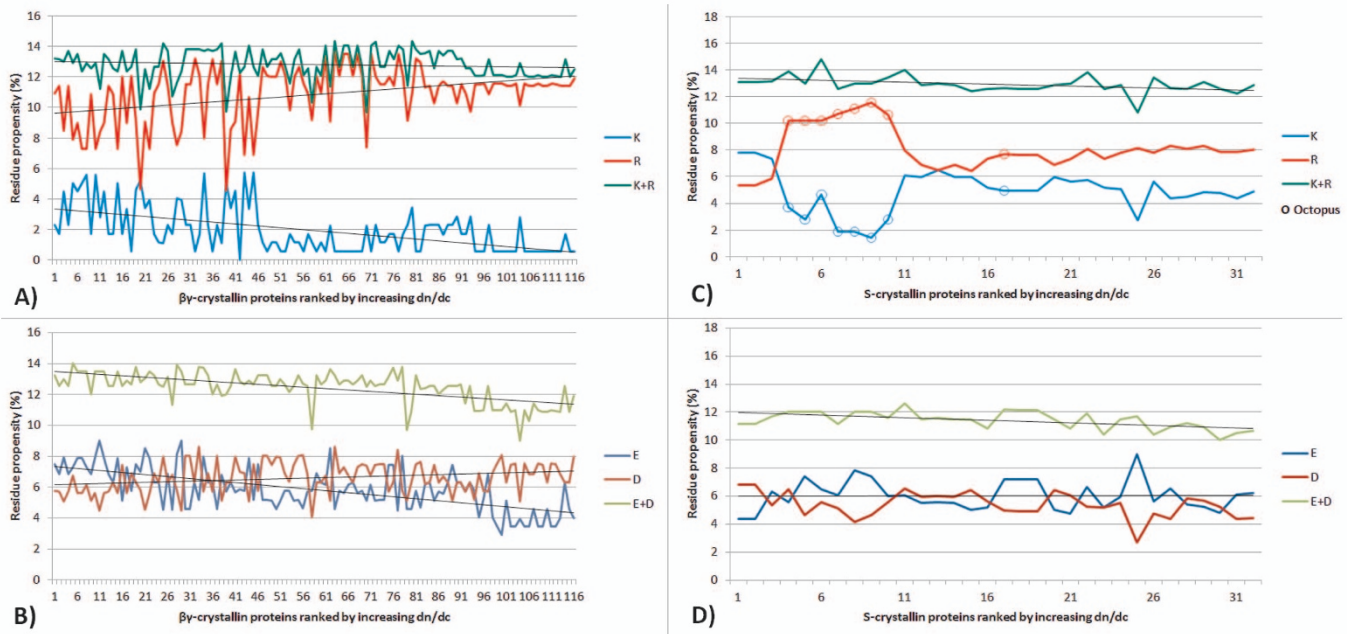


Figure 4 | The number of residues plotted for each respective protein from the selected datasets of $\beta\gamma$ - and S-crystallins, showing in A) and C) proportion of Lysine (K) and Arginine (R) residues and in B) and D) the number of Glutamic Acid (E) and Aspartic Acid (D) residues. $\beta\gamma$ -crystallin K-R correlation: $-0.85, p < 0.0005$; $\beta\gamma$ -crystallin E-D correlation: $-0.70, p < 0.0005$; S-crystallin K-R correlation $-0.92, p < 0.0005$; S-crystallin E-D correlation: $-0.81, p < 0.0005$. Open circles correspond to S-crystallins from octopus.

ilar trend: that the double Greek key motif in crystallins has a higher dn/dc than those in other proteins. A Greek key motif is therefore not necessarily indicative of a high dn/dc value. Sequence length in Greek keys was not found to be correlated with dn/dc. The dn/dc values of the β -hairpins in the Greek keys from isoforms γB , $\beta B2$, $\beta B3$, $\beta B1$ and $\beta A4$ are slightly higher than the dn/dc values of the whole sequence. Whether the amino acids in these structural regions have a dual role in contributing to refractive index and to structural stability requires further investigation.

Tighter packing of proteins will increase the refractive index as proteins have a higher refractive index than water. The substitutions of arginine for lysine and aspartic acid for glutamic acid, not only increase the dn/dc value but could also result in a more compact protein structure. The guanidinium group on the side chain of arginine has a geometry and charge distribution that renders it able to form multiple hydrogen bonds; the shorter, less flexible side chain of aspartic acid compared to glutamic acid may also facilitate compaction.

Both in the $\beta\gamma$ -crystallins and in the S-crystallins, the higher the dn/dc value of a protein, the fewer lysine and glutamic acid residues and the more arginine and aspartic acid residues it contains. Since lysine/arginine and glutamic/aspartic acid residues are associated with the formation of salt bridges³⁸, the observed substitutions may

be constrained by the need to maintain existing salt bridges. Analysis of representative $\beta\gamma$ crystallin protein structures revealed that mammalian $\beta\gamma$ -crystallins have a much higher proportion of salt bridges per domain than would be expected given the domain length of about 90 residues (5 salt bridges compared to a standard of <2 salt bridges for a domain of that length³⁸). This pattern is not found among the three non-mammalian crystallins investigated in this study; these display only one salt bridge per domain on average. It is notable that salt bridges are more frequent in α -helical structures³⁸, whereas $\beta\gamma$ -crystallins have a relatively high proportion of β -pleated sheet.

A cross species comparison showed that arginine consistently replaced lysine with progression from lower to higher dn/dc value proteins within a species. The glutamic acid/aspartic acid correlation was not borne out in this comparison. Instead another trend was observed: a decrease in phenylalanine and a concomitant increase in tyrosine with increase in dn/dc value. As both phenylalanine and tyrosine have relatively high dn/dc values, (0.244 and 0.240 ml/g respectively) which are very close in magnitude, such a substitution makes little difference to the refractive index.

The findings of this study show relatively higher numbers of salt bridges in mammalian $\beta\gamma$ -crystallins than in other proteins³⁸ and there is a predominance of salt bridges formed between amino acids separated by more than 10 residues. Additionally, the relatively high

Table 1 | Refractive index increment of selected sequence

	XENLA	DANRE	MOUSE	BOVINE	HUMAN
Beta-B2 crystallin					
(reference)	(Q6DJC7)	(Q52J14)	(P62696)	(P02522)	(P43320)
Refractive increment	0.1958	0.1950	0.1942	0.1939	0.1942
Gamma crystallin with low dn/dc					
(reference)	Gamma-E-crystallin (Q6DJC9)	Gamma M5-crystallin (Q5XJ63)	Gamma-S-crystallin (O35486)	Gamma-S-crystallin (P06504)	Gamma-S-crystallin (P22914)
Refractive increment	0.2001	0.1974	0.1981	0.1988	0.1984
Gamma crystallin with high dn/dc					
(reference)	Gamma-A-crystallin (Q66KU7)	Gamma M2-crystallin (A7E2K8)	Gamma-crystallin E (Q03740)	Gamma-F-crystallin (P23005)	Gamma-A-crystallin (P11844)
Refractive increment	0.2074	0.2077	0.2033	0.2021	0.1999



Table 2 | Interspecies comparison of amino acid substitutions

Species	Low- > High γ				β B2 -> Low γ				β B2 -> High γ			
	K->R	R->K	F->Y	Y->F	K->R	R->K	F->Y	Y->F	K->R	R->K	F->Y	Y->F
XENLA	6	0	6	0	2	1	1	1	3	0	2	1
DANRE	5	0	6	0	3	1	1	2	4	0	2	0
MOUSE	3	0	2	0	4	0	3	0	3	1	2	1
BOVINE	4	0	3	0	3	1	1	1	4	0	3	0
HUMAN	4	0	3	0	2	1	2	1	4	0	3	0

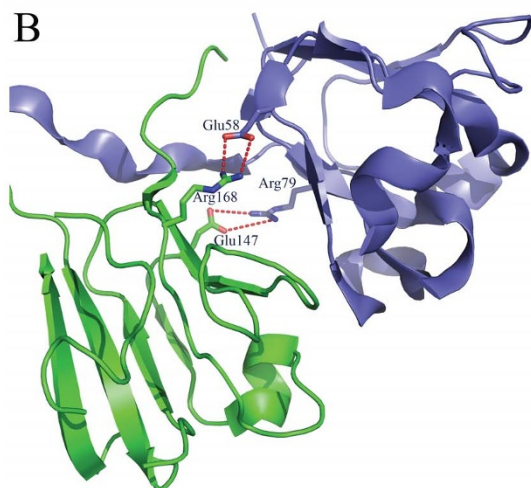
proportion of interstrand bridges in the crystallins compared to what has been found in other proteins^{38,39} may be indicative of long range structural stability. This is of particular importance to the crystallins which remain in the cytoplasm of lens fibre cells from their synthesis to death of the organism. In the case of cells from the central regions of lenses, protein synthesis has taken place during gestation and maintenance of optical quality is required for decades. The highest content of γ -crystallins is found in the central regions of mammalian lenses where refractive index reaches maximum magnitude^{3,16,17}; in cephalopods the predominant proteins which contribute to the high refractive index are S-crystallins^{3,14,15}. Both protein classes have relatively high refractive increments.

Whilst the optical function of the lens is to provide refractive power, the quality of the optics relies on transparency. Cataract results in a loss of transparency and it has recently been shown that congenital mutations in human γ D-crystallin can cause cataract to develop with or without disruption to the Greek key structure⁴⁰. Mutations such as that which results in substitution of arginine by serine at position 77 (R77S) do not destabilise the tertiary structure yet result in cataract in the cortical regions of the lens⁴⁰. Other mutations such as the one that leads to substitution of proline for alanine (A36P), disrupt the Greek key structure and cause nuclear cataract⁴⁰. Single substitutions similar to the aforementioned do not produce any substantive change in the value of dn/dc .

A

pdb id		2JDF		1BLB		3QK3		1OKI		3LWK	
γ B	D1	D21	R59	P21	E58	E44	E82	E78	E116	E70	C99
β B2	D1 A/B										
β B3	D1										
β B1	D1										
β A4	D1										
pdb id		2JDF		1BLB		3QK3		1OKI		3LWK	
γ B	D2	D108	R147	D108	E147	D135	E174	D169	E208	E165	C166
β B2	D2 B/A										
β B3	D2										
β B1	D2										
β A4	D2										
			R168		R168		R195		R229		R192
			R169		R169		R196		R230		R193
			M171		R171		R198		-		Q195

B



C

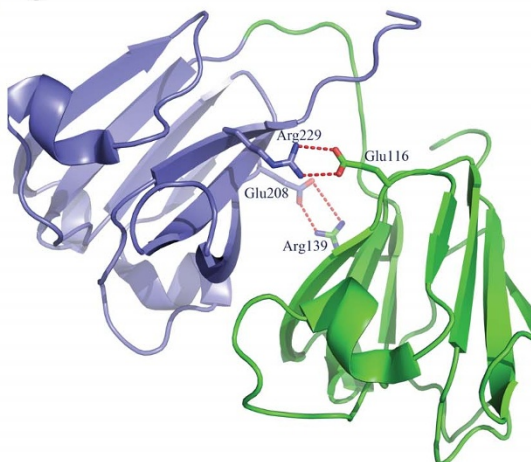


Figure 5 | Salt bridge conservation in $\beta\gamma$ -crystallins. (A) Multiple alignment of $\beta\gamma$ -crystallin sequences including alignment of the two domains. Only residues relevant to conserved salt bridges and disulphide bond are displayed. Green and red numbers represent negatively and positively charged residues, respectively. Yellow dashed lines show disulphide bonds. Pale blue shaded residues are involved in cross domain salt bridges. Navy shaded residues are involved in completely conserved salt bridges⁵¹; (B) inter protein disulphide bridges observed in a homodimer of crystallin β B2. Monomers are shown in different colours⁵¹; (C) inter domain salt bridges formed within a single crystallin β B1 monomer. Salt bridge interactions are shown as red dotted lines⁵².



The lenticular crystallins are organised to ensure that the lens meets the refractive demands of the eye. The $\beta\gamma$ -crystallins and the S-crystallins contain residues that contribute to a high dn/dc when compared to non-lenticular proteins and the Greek key motif which exists in many proteins, is linked with a higher dn/dc only when it is found in crystallins. Salt bridge interactions that stabilise protein structure and provide interactive potential, are relevant to structural longevity of the crystallins and are necessary for maintenance of transparency over decades. The crystallins have not only evolved with a primary sequence that optimises their contribution to refraction, they have higher order arrangements that are conducive to its preservation.

Methods

Analysis of $\beta\gamma$ -crystallin refractive index increment. Refractive index increments (for 589 nm at 25°C) were predicted for all available sequences belonging to the $\beta\gamma$ -crystallin family. Those sequences were retrieved using relevant seed sequences as Psi-Blast queries⁴¹. In order to ensure the widest coverage while avoiding the introduction of unrelated proteins, seeds were defined as all sequences whose annotations, based on experimental data, specify the molecular function as “the action of a molecule that contributes to the structural integrity of the lens of an eye” (GO:0005212⁴²) and identify them as belonging to the $\beta\gamma$ -crystallin family. As results, 13 seed sequences were used: 7 from mouse - β -crystallin A1/A2/B2/S and γ -crystallin B/C/D/E - and 6 from rat - β -crystallin A4/B1/B3 and γ -crystallin C/D/E.

All sequences returned by Psi-Blast with an e-value below 1 were mapped to UniRef100⁴³ to remove duplicates and fragments. Since the $\beta\gamma$ -crystallin superfamily contains a few non-crystallin members, such as absent in melanoma 1 (AIM1)⁴⁴, their associated sequences were removed from the initial list. This was performed by generating a phylogenetic tree using FastTree⁴⁵ from a multiple alignment⁴⁶ and eliminating sequences belonging to non-crystallin branches. Eventually, the diversity of the $\beta\gamma$ -crystallin family was represented by 292 entries. Prediction of their dn/dc values was performed following the computational method outlined by McMeekin et al.²⁶ and described in previous studies^{14,15}.

Analysis of S-crystallin refractive index increment. Refractive index increments (for 589 nm at 25°C) were predicted for relevant sequences belonging to the S-crystallin family. Those sequences were retrieved using a seed sequence - squid S-crystallin (P18426) - as Psi-Blast query⁴⁰. All sequences returned with an e-value below 1 were candidates for further filtering. Sequences not belonging to the S-crystallin family such as its homologous glutathione S-transferases were discarded. Fragments, predicted and hypothetical sequences were also not considered. Finally, among the 38 remaining S-crystallin sequences, only those belonging to cephalopod species were selected for our study, ie 24 from Squid and 8 from Octopus; 6 Oyster sequences were discarded.

Analysis of Greek key sequences. Most $\beta\gamma$ -crystallins contain a double Greek key motif that has a length of 170–180 residues. To study the constraint this motif confers on protein evolution, sequences outside that length range were removed from the dataset. Hence, 116 $\beta\gamma$ -crystallin sequences were used for analysis. Predicted dn/dc values of $\beta\gamma$ -crystallins were compared to those of a set of 52 double Greek key domain representatives²⁷, whose sequences were retrieved from the Protein Data Bank⁴⁷.

Salt bridge analysis. Salt bridge analysis was performed on a set of 3D protein structures that are representative of $\beta\gamma$ -crystallins. These structures were extracted from the Protein Data Bank using a 90% sequence similarity filter to eliminate duplicates and non-wild type copies. In addition to non-crystallin members, two AIM1 proteins and a crystallin whose domains were artificially permuted⁴⁸ were also removed from the list. A total of 19 structures were studied: 7 human, 5 cow, 3 mouse, 1 rat, 1 sea squirt, 1 bacterium and 1 archaea. Descriptions of salt bridges were computed by the Salt Bridges Plugin of the molecular graphics program, VMD⁴⁹; their classification as interdomain and interstrand and calculation of β -hairpins was performed using descriptions produced by PROMOTIF⁵⁰.

1. Nilsson, D. E. Eye evolution and its functional basis. *Vis Neurosci* **30**, 5–20 (2013).
2. Nilsson, D. E. The evolution of eyes and visually guided behaviour. *Philos Trans R Soc Lond B Biol Sci* **364**, 2833–2847 (2009).
3. Pierscionek, B. K. & Regini, J. W. The gradient index lens of the eye: an opto-biological synchrony. *Prog Retin Eye Res* **31**, 332–349 (2012).
4. Song, S. et al. Functions of the intermediate filament cytoskeleton in the eye lens. *J Clin Invest* **119**, 1837–1848 (2009).
5. Clark, A. R., Lubsen, N. H. & Slingsby, C. sHSP in the eye lens: crystallin mutations, cataract and proteostasis. *Int J Biochem Cell Biol* **44**, 1687–1697 (2012).
6. Delaye, M. & Tardieu, A. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature* **302**, 415–417 (1983).
7. Wistow, G. J. & Piatigorsky, J. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu Rev Biochem* **57**, 479–504 (1988).

8. Bloemendal, H. et al. Ageing and vision: structure, stability and function of lens crystallins. *Prog Biophys Mol Biol* **86**, 407–485 (2004).
9. Bassnett, S., Shi, Y. & Vrensen, G. F. Biological glass: structural determinants of eye lens transparency. *Philos Trans R Soc Lond B Biol Sci* **366**, 1250–1264 (2011).
10. Kroger, R. H., Campbell, M. C. & Fernald, R. D. The development of the crystalline lens is sensitive to visual input in the African cichlid fish, *Haplochromis burtoni*. *Vision Res* **41**, 549–559 (2001).
11. Hoshino, M. et al. Optical properties of in situ eye lenses measured with X-ray Talbot interferometry: a novel measure of growth processes. *PLoS One* **6**, e25140 (2011).
12. Barer, R. & Joseph, R. Refractometry of living cells. *Quart. J. Microscop. Sci* **95**, 399–423 (1954).
13. Pierscionek, B., Smith, G. & Augusteyn, R. C. The refractive increments of bovine alpha-, beta-, and gamma-crystallins. *Vision Res* **27**, 1539–154 (1987).
14. Zhao, H., Brown, P. H., Magone, M. T. & Schuck, P. The molecular refractive function of lens gamma-crystallins. *J Mol Biol* **411**, 680–699 (2011).
15. Zhao, H., Brown, P. H. & Schuck, P. On the distribution of protein refractive index increments. *Biophys J* **100**, 2309–2317 (2011).
16. Keenan, J., Orr, D. F. & Pierscionek, B. K. Patterns of crystallin distribution in porcine eye lenses. *Mol Vis* **14**, 1245–1253 (2008).
17. Keenan, J., Elia, G., Dunn, M. J., Orr, D. F. & Pierscionek, B. K. Crystallin distribution patterns in concentric layers from toad eye lenses. *Proteomics* **9**, 5340–5349 (2009).
18. Pierscionek, B. K. & Chan, D. Y. Refractive index gradient of human lenses. *Optom Vis Sci* **66**, 822–829 (1989).
19. Pierscionek, B. K. Refractive index contours in the human lens. *Exp Eye Res* **64**, 887–893 (1997).
20. Tardieu, A., Veretout, F., Krop, B. & Slingsby, C. Protein interactions in the calf eye lens: interactions between beta-crystallins are repulsive whereas in gamma-crystallins they are attractive. *Eur Biophys J* **21**, 1–12 (1992).
21. Jaenicke, R. Stability and folding of domain proteins. *Prog Biophys Mol Biol* **71**, 155–241 (1999).
22. Purkiss, A. G. et al. Biophysical properties of gamma C-crystallin in human and mouse eye lens: the role of molecular dipoles. *J Mol Biol* **372**, 205–222 (2007).
23. Dorsaz, N. & Foffi, G. Phase behaviour of the symmetric binary mixture from thermodynamic perturbation theory. *J Phys Condens Matter* **22**, 104113 (2010).
24. Blundell, T. et al. The molecular structure and stability of the eye lens: X-ray analysis of gamma-crystallin II. *Nature* **289**, 771–777 (1981).
25. Aravind, P., Mishra, A., Suman, S. K., Jobby, M. K., Sankaranarayanan, R. & Sharma, Y. The betagamma-crystallin superfamily contains a universal motif for binding calcium. *Biochemistry* **48**, 12180–12190 (2009).
26. McMeekin, T. L., Groves Merton, L. & Hipp Norbert, J. in *Amino Acids and Serum Proteins* Vol. 44 *Advances in Chemistry* Ch. 4, 54–66 American Chemical Society, 1964).
27. Zhang, C. & Kim, S. H. A comprehensive analysis of the Greek key motifs in protein beta-barrels and beta-sandwiches. *Protein* **40**, 409–419 (2000).
28. Hand, D. B. The refractivity of protein solutions. *J Biol Chem* **108**, 703–707 (1935).
29. Perlmann, G. & Longman, L. G. The Specific Refractive Increment of Some Purified Proteins. *J. Am. Chem. Soc.* **8**, 2719–2724 (1948).
30. Bjork, I. Studies on gamma-crystallin from calf lens. I. Isolation by gel filtration. *Exp Eye Res* **1**, 145–154 (1961).
31. Andries, C. et al. Physical-chemical studies on bovine eye lens proteins. I. Light-scattering and viscosity studies of low-molecular weight alpha-crystallin isolated from adult and embryonic bovine lenses. *Exp Eye Res* **34**, 239–255 (1982).
32. Perry, A. J. & Koenig, V. L. Some physicochemical properties of the water-soluble proteins of the bovine eye lens. *Biochim Biophys Acta* **46**, 413–422 (1961).
33. Harding, J. J. & Dilley, K. J. Structural proteins of the mammalian lens: a review with emphasis on changes in development, aging and cataract. *Exp Eye Res* **22**, 1–73 (1976).
34. Bloemendal, H. Lens proteins. *CRC Crit Rev Biochem* **12**, 1–38 (1982).
35. Srikanthan, D., Bateman, O. A., Purkiss, A. G. & Slingsby, C. Sulfur in human crystallins. *Exp Eye Res* **79**, 823–831 (2004).
36. Bjork, I. Studies on gamma-crystallin from calf lens. 3. Comparison of the main protein components by peptide mapping. *Exp Eye Res* **9**, 152–157 (1970).
37. Croft, L. R. The amino acid sequence of γ -crystallin (fraction II) from calf lens. *Biochem J* **128**, 961–970 (1972).
38. Sarakatsannis, J. N. & Duan, Y. Statistical characterization of salt bridges in proteins. *Proteins* **60**, 732–739 (2005).
39. Donald, J. E., Kulp, D. W. & DeGrado, W. F. Salt bridges: geometrically specific, designable interactions. *Proteins* **79**, 898–915 (2011).
40. Vendra, V. P. et al. Structural integrity of the Greek key motif in $\beta\gamma$ -crystallins is vital for central eye lens transparency. *PLoS ONE* **8**, e70336 (2013).
41. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).
42. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29 (2000).
43. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
44. Wistow, G. The human crystallin gene families. *Hum Genomics* **6**, 26 (2012).
45. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).



46. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D. & Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
47. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res* **28**, 235–242 (2000).
48. Wright, G., Basak, A. K., Wieligmann, K., Mayr, E. M. & Slingsby, C. Circular permutation of betaB2-crystallin changes the hierarchy of domain assembly. *Protein Sci* **7**, 1280–1285 (1998).
49. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33–38, 27–38 (1996).
50. Hutchinson, E. G. & Thornton, J. M. PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212–220 (1996).
51. Nalini, V. *et al.* Close packing of an oligomeric eye lens beta-crystallin induces loss of symmetry and ordering of sequence extensions. *J Mol Biol* **236**, 1250–1258 (1994).
52. Van Montfort, R. L., Bateman, O. A., Lubsen, N. H. & Slingsby, C. Crystal structure of truncated human betaB1-crystallin. *Protein Sci* **12**, 2606–2612 (2003).

Author contributions

J.C.N. and B.P. conceived the idea; K.M., C.E. collected data, J.C.N., K.M., B.P. and A.R. analysed the data, J.C.N., B.P. wrote the first draft of the manuscript; J.C.N., K.M. B.P. and A.R. contributed to writing the final version of the manuscript; J.C.N. and A.R. produced the figures.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mahendiran, K., Elie, C., Nebel, J.-C., Ryan, A. & Pierscionek, B.K. Primary sequence contribution to the optical function of the eye lens. *Sci. Rep.* **4**, 5195; DOI:10.1038/srep05195 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images in this article are included in the article's Creative Commons license, unless indicated otherwise in the image credit; if the image is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the image. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>